JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1998, **31,** 431–446 NUMBER 3 (FALL 1998)

IDENTIFICATION AND VALIDATION OF CRITICAL CONVERSATIONAL SOCIAL SKILLS

CAROLYN HUGHES, SARAH W. LORDEN, STACEY V. SCOTT, BOGSEON HWANG, KRISTINE R. DERER, MICHAEL S. RODI, SARAH E. PITKIN, AND JOY C. GODSHALL

VANDERBILT UNIVERSITY

In this study, we used a four-step social validation process to identify and validate critical skill components that constitute high school students' conversational behavior. The four steps were nominating target behaviors, establishing a normative range of performance, manipulating simulations of behavioral dimensions, and comparing ratings of judges to levels of performance on those behavioral dimensions. Multiple measures, both quantitative and qualitative, suggested that the rate and percentage of time initiating and responding verbally, the percentage of time attending, and the percentage of time not engaging in distracting motor behavior related to favorable ratings by a wide variety of 60 judges. Findings are discussed in relation to the utility of the multistep social validation process and the identification of critical social skill components as targets of interventions.

DESCRIPTORS: social validation, high school students, mental retardation, conversational social skills

A successful transition from school to adult life requires the performance of myriad skills, such as getting along with coworkers, fixing one's meals, and paying one's bills (Clark, Field, Patton, Brolin, & Sitlington, 1994; Halpern, 1992; Hughes, Hwang, Kim, Eisenman, & Killian, 1995; Schalock, 1986). Because instructional time available to teach such skills to students with disabilities is limited, it is important to identify critical component behaviors that relate to socially valued outcomes such as maintaining a job, living a healthy lifestyle, and being a good citizen (Brown et al., 1979; Moore, Agran, & McSweyn, 1990; Rusch, DeStefano, Chadsey-Rusch, Phelps, & Szymanski, 1992). Social skills frequently are mentioned as critical to outcomes such as social interaction and social relationships (Siperstein, 1992). However, few empirical investigations have identified essential skill components that relate to favorable social outcomes (Butterworth & Strauch, 1994; Chadsey-Rusch, 1992; Hughes, Killian, & Fischer, 1996).

Rather than choosing to modify social behaviors selected on the basis of face validity alone, as derived from the literature or the judgment of an investigator, researchers have argued for the empirical and social validation of such skills. The validation process can insure that instructional time is invested in modifying behavior that is relevant to an individual's everyday life (Goldstein, Kaczmarek, Pennington, & Shafer, 1992; Nelson & Hayes, 1979; Schwartz & Baer, 1991; Van Houten, 1979; Wolf, 1978). For example, the use of social amenities may not be practiced or valued in some work environments (Ferguson, McDonnell, & Drew,

This research was supported in part by Grant H023N10017 from the Office of Special Education and Rehabilitative Services, U.S. Department of Education and the John F. Kennedy Center for Research on Human Development at Vanderbilt University.

We thank the students and teachers at McGavock High School, Nashville, Tennessee, and 60 anonymous raters for their participation and assistance in conducting this study. We also thank Lotte Whatley Shipman and Dana Patterson for acting in the video scenes.

Correspondence concerning this article should be addressed to Carolyn Hughes, Department of Special Education, Box 328 Peabody, Vanderbilt University, Nashville, Tennessee 37203 (E-mail: carolyn.hughes@vanderbilt.edu).

1993) or joking with coworkers on the job may be more appropriate in one setting (e.g., break room) than in another (e.g., work station; Chadsey-Rusch, Gonzalez, Tines, & Johnson, 1989). Further, researchers should identify the optimal frequency of performance of social skills (e.g., initiating) that produces maximal desired outcomes, such as reciprocal conversation or social acceptance (Goldstein et al., 1992; Hawkins, 1991; Warren, Rogers-Warren, & Baer, 1976).

One means of identifying and validating critical target behaviors and their components is a four-step process that we have derived from the work of Hawkins (1991), Minkin et al. (1976), Van Houten (1979), and others. First, researchers should observe social interaction in a targeted environment to nominate component skills that constitute a behavior of interest (Bakeman & Gottman, 1986; Goldstein et al., 1992; Minkin et al., 1976). Second, investigators should develop reliable definitions of the component skills and establish a normative range of performance within the environment (Nelson & Hayes, 1979). Third, investigators should experimentally manipulate performance of the component skills across a range of relevant dimensions (e.g., frequency or duration; Hawkins, 1991). Fourth, a variety of relevant judges (e.g., peers, community members, parents) should evaluate the range of behaviors (Minkin et al., 1976). If specific ranges of performance (e.g., 90% time spent attending to partner while engaged in conversation) correlate with favorable ratings by judges, these targets should be incorporated into instructional curricula as potentially functional goals (Fawcett, 1991; Van Houten, 1979). Despite the value of empirically and socially validating an optimal range of target behaviors prior to instruction, few researchers have conducted such a four-step process (Hawkins, 1991).

Two recent studies reported efforts to validate components of social behavior that may be relevant to social skills interventions for adults with mental retardation (Quinn, Sherman, Sheldon, Quinn, & Harchik, 1992; Sherman, Sheldon, Harchik, Edwards, & Quinn, 1992), using procedures similar to the work of Minkin et al. (1976). In both studies, the investigators assessed the social validity of behaviors that frequently are targets of social skills interventions (i.e., following instructions, accepting criticism, resolving conflicts) by comparing levels of performance of target behaviors with judges' ratings of the acceptability of performance. Although judges' ratings correlated with the performance of the three social behaviors, target behaviors were not empirically derived, as recommended by Van Houten (1979) and others. Instead, target behaviors were selected on the basis of face validity alone as derived from the literature.

The purpose of this study was to identify and validate critical skill components of high school students' conversational behavior that could serve as instructional targets for teaching secondary students with mental retardation. We sought to validate relevant skills empirically and socially by conducting a four-step process that consisted of (a) directly observing conversational interaction in the actual environment (i.e., high school lunchroom) to nominate target skill components, (b) reliably defining the skill components and establishing a normative range of performance, (c) experimentally manipulating simulated performances of the skill components across potentially relevant dimensions, and (d) obtaining judges' ratings of performance at various levels of those dimensions. Unlike Quinn et al. (1992) and Sherman et al. (1992), we investigated an everyday social skill (i.e., conversing with peers in a high school lunchroom) rather than task-related social skills (e.g., following instructions), and we focused on high school students rather than on adults. Further, we used social comparison methods (Kazdin,

1977) to identify component skills rather than selecting behavior targets based on face validity. Finally, unlike the previous studies, we assessed the perspectives of peers and students who had mental retardation regarding the acceptability of the targeted social performance.

METHOD

Nomination of Target Behaviors

To identify target behaviors, we first conducted direct observations of students' conversational interactions in the lunchroom of a large, urban high school that enrolled 2,700 students and offered courses in academic and vocational preparation. The student population was 58% Caucasian; 40% African American; and 2% Asian American, Native American, and other ethnic groups. The high school served general education students as well as students with disabilities (e.g., mental retardation, autism, learning or behavior disorders; $N = 325$). Students with disabilities were served in partially mainstreamed, self-contained, or resource room classes. All general and special education students ate lunch in unassigned seating in one of two large lunchrooms, in which all observations of students' conversational behavior were conducted.

Preliminary observation indicated that although general and special education students ate in proximity of each other, virtually no social interaction occurred between members of the two groups. This observation prompted us to investigate how interactions of the two groups differed, reasoning that if students with disabilities displayed social behaviors similar to those of their general education peers, social interaction between individuals in the two groups might increase (Haring, 1991). Therefore, we chose to nominate potential target behaviors by informally observing a total of 320 general education students and 39 students with moderate or severe mental retardation. Twice weekly for 2 months, we observed and recorded sequentially in narrative form all physical and verbal behaviors of the students and the social context in which the behaviors occurred, using the students' naturally occurring lunch groups as the unit of analysis. From this observation, we identified 97 behaviors and 33 conversational topics that occurred during conversational interaction. Three investigators independently grouped the listed behaviors and topics by their apparent function into conversational interaction categories that emerged as the investigators examined the data (e.g., attending to social focal point or talking about peers). The three investigators then compared their lists, and interrater reliability was calculated via point-by-point agreement (Kazdin, 1982) for each category. An agreement was scored only if all three investigators placed a behavior or topic in the same category. Overall interrater reliability averaged 83% (range across categories), which was computed by dividing agreements by agreements plus disagreements and multiplying by 100%. The process resulted in a list of nine conversational behavior categories and 10 conversational topic codes.

We then applied those nine behavior and 10 topic codes in observing two groups of 12 students in the same setting. Students in both groups represented the age, gender, and ethnic composition of the high school's student population. Students selected for one group had moderate ($n = 10$) or severe ($n = 2$) mental retardation and functional communication skills. Twelve general education students who were judged by the authors' observation to be socially competent (e.g., they frequently exchanged social bids with their peers) and to engage in conversation often were chosen for the second group. This group's selection criteria were designed to insure that target behaviors were performed at a level considered normative

and functional within the setting as a means of social comparison with the students with mental retardation (Kazdin, 1977; Van Houten, 1979). By comparing the two groups, we sought to identify behaviors that discriminated between students with and without mental retardation that could serve as potential instructional targets (Hawkins, 1991).

For 3 months, we conducted 10 observations of each of the 24 students. Total observation time per student averaged 2 hr 50 min. As in previous investigations of social interaction (e.g., Storey & Horner, 1991), the students were aware of the observations but were unaware of the behaviors of interest. We recorded data using laptop computers equipped with a Multiple Option Observation System for Experimental Studies (MOOSES) software program (Tapp, Wehby, & Ellis, 1995). MOOSES allows simultaneous recording of event and duration based in real time. In addition, the percentage of time that the behaviors occurred was computed as the cumulative duration of each behavior. We also recorded the context in which the behaviors occurred (e.g., number, gender, and ethnicity of peers; teacher or staff interaction) and the content of topics discussed. Interobserver agreement was taken during 30% of all observation sessions per behavior per student and was calculated by using the point-by-point agreement method (Kazdin, 1982). Mean overall agreement per behavior ranged from 90% to 99%.

Findings revealed significant mean differences between the groups in four of the nine behavioral categories: (a) rate and percentage of time initiating, (b) rate and percentage of time responding, (c) percentage of time attending to the focal person or social focal point, and (d) percentage of time engaging in distracting motor behavior. (See definitions of behaviors in Table 1 and mean differences in Table 2. Percentage of time was the cumulative duration of a behavior calculated as a percentage of total observational time.) Significant differences in frequency were also observed for the 10 conversational topics discussed by the two groups of students: peers; food; school events (social); afterschool events; jokes; school events (academic); money; television, movies, bands, and celebrities; work and employment; and repetitive topics or nondiscernible speech. Briefly, general education students discussed peers, school happenings, and social events and told jokes much more frequently than did students with mental retardation.

In contrast, the social context in which interactions occurred was strikingly similar for the two groups of students. Similarities observed between the groups were as follows: (a) All students overwhelmingly displayed a positive or neutral affect (e.g., smiling, maintaining a relaxed body position, making positive remarks) rather than negative affect (e.g., frowning, maintaining poor body posture, complaining) while interacting with peers; (b) interactions between students with and without mental retardation were virtually nonexistent; (c) interactions between students and teachers or staff were negligible; and (d) naturally occurring groups of students were similar in number, gender, ethnicity, and seating arrangement. In addition, initiations or responses judged to be inappropriate within the social context occurred less than 0.03 times per minute for students with or without mental retardation (interobserver agreement = 92%, range, 73% to 100%). (An initiation or response was considered inappropriate if the volume, tone or quality of voice, pitch, intensity, intonation, or rate, time of occurrence, or topic of the utterance was not consistent with standards established by social comparison within the lunchroom or if it produced a negative effect on a conversational partner.) In summary, the primary differences that we observed between students with and without mental retardation were in rate and percent-

Table 1
Definitions of Behaviors with Significant Differences Between Two Groups of Students

Behavior	Definition
Initiating	Verbal or nonverbal behavior directed toward another person that introduces a new topic or expansion of an existing topic, introduces new information not related to information from a prior utterance, or is preceded by at least 15 s with no interactive verbal behavior with the same person. Includes communicative gestures such as waving (Fey, 1986; Foster & Cone, 1986).
Responding	Verbal or nonverbal behavior in response to an initiation without expanding on a topic or adding new information to a prior utterance. Includes asking for clarification of an initiation and meaningful nonword verbalizations or gestures that serve as acknowledgments or responses, such as "hm-m-m," "uh-huh," shaking head "yes" or "no," smiling, frowning, waving in response, pointing, winking, or shrugging shoulders (Breen, Kennedy, & Haring, 1991; Fey, 1986).
Attending to focal person or social focal point	Participant attends and shifts attention appropriately and promptly to relevant social stimuli in the immediate environment as indicated by directing face toward social focal point (e.g., participant sitting with a group of peers at a table shifts attention as speakers shift during conversation). Attending is not scored if participant does not attend to or shift attention appropriately and promptly to relevant social stimuli in the immediate environment as indicated by not directing face toward social focal point (e.g., participant holds fixed gaze away from speaker at table; Koegel & Frea, 1993).
Engaging in distracting motor behavior	Participant performs a motor behavior that would be considered socially inappropriate when compared to the behavior of peers within the immediate environment or that an observer judges to be interfering in the occurrence of social interaction between the participant and peers or teachers (e.g., continuously rocking torso back and forth, covering face with hands, hitting own chin with hand; Koegel & Frea, 1993).

age of time initiating, responding, attending, and engaging in distracting motor behavior rather than in the social context or conditions within which interactions occurred (e.g, when an initiation occurred during a conversation, the content of what was discussed, or the quality of a speaker's voice).

Establishment of a Normative Range of Performance

We used measures of mean performance for the group of 12 general education students to derive a normative range of conversational interaction for high school students while they eat lunch at school. Using a convention derived from the field of language disorders (Prutting & Kirchner, 1987), we defined a normative range as one standard deviation (*SD*) above and below the mean performance of the general education students (see Table 2). (Note that because the standard deviations for the target behaviors were small relative to their means, the normative ranges established for the target behaviors were comparatively narrow.)

Manipulation of Behavioral Dimensions

To socially validate the identified performance, we experimentally manipulated levels of the six target behaviors shown in Table 2, as performed by two amateur actors in four videotaped scenes. To minimize effects of confounding variables, we held constant all the contextual variables we had observed among general education students (e.g., affect, conversational topics, appropriateness of interactions). In addition, the effects of actors' characteristics (e.g., voice quality or facial tic) were held constant by having judges rate the performance of only one actor across videos. Therefore, differences in ratings could be attributed to manipulations of

Table 2
Means, Standard Deviations, and Normative Range of Target Behaviors

Behavior	Students with mental retardation		Students without mental retardation		Normative range	*t*
	M	*SD*	*M*	*SD*		
Rate[a] of initiating	0.9	0.8	3.2	0.5	2.7–3.7	8.58*
Percentage of time[b] initiating	4.8	4.2	18.0	2.7	15.3–20.7	9.12*
Rate of responding	0.5	0.4	2.2	0.4	1.8–2.6	10.10*
Percentage of time responding	2.1	1.7	6.9	1.4	5.5–8.3	7.55*
Percentage of time attending to focal person or focal social point	56.4	18.6	89.0	4.0	85.0–93.0	5.97*
Percentage of time engaging in distracting motor behavior	24.1	19.1	0.5	0.5	0–1.0	−4.28*

[a] Per minute.
[b] Percentage of total observation time.
* *t* test, $df = 22$, $p < .0001$.

the target behaviors rather than an actor's idiosyncratic behavior.

The actors used scripts that systematically varied rate and percentage of time performing the target behaviors. This systematic variation is portrayed in Table 3. Behaviors were performed either within, above, or below the normative range established for the general education students. Behaviors performed within the normative range were

Table 3
Manipulation of Behavioral Dimensions on Videos and Means, Standard Deviations, and MANOVA Analysis of Judges' Responses

Behavior	Manipulation of behavioral dimensions			
	Video 1	Video 2	Video 3	Video 4
Rate of initiating	Within[a]	Below[b]	Within	Below
Percentage of time initiating	Within	Below	Within	Below
Rate of responding	Within	Below	Within	Below
Percentage of time responding	Within	Below	Within	Below
Percentage of time attending to focal person or focal social point	Within	Within	Below	Below
Percentage of time engaging in distracting motor behavior	Within	Within	Within	Above[c]

Means, standard deviations, and MANOVA analysis of judges' responses to quantitative questions[d]

	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1. The student has good conversational skills.	4.7	0.7	1.4	0.8	3.3	1.3	1.2	0.4
2. The student acts like most high school students when they eat lunch in a school cafeteria.	4.6	0.9	1.6	1.0	3.6	1.1	1.5	0.8
3. Most high school students would probably act like the student in this conversation.	4.4	0.9	1.4	0.8	3.4	1.2	1.4	0.6
4. The student's conversational behavior looks acceptable to me.	4.6	0.7	1.6	0.9	2.8	1.4	1.3	0.6
5. Most high school students would probably enjoy having a conversation with someone who acts like the student.	4.7	0.6	1.5	0.9	3.2	1.2	1.3	0.6
Mean score	4.6	0.6	1.5	0.8	3.2	1.0	1.3	0.4

Note. MANOVA for repeated measures, $F = 315.99$, $df = 3, 177$, $p < .0001$.
[a] Within normative range.
[b] Below normative range.
[c] Above normative range.
[d] 1 = *strongly disagree,* 2 = *mildly disagree,* 3 = *do not feel strongly either way,* 4 = *mildly agree,* 5 = *strongly agree.*

performed at less than one standard deviation of the mean performance of students without mental retardation. Behaviors above or below the normative range were performed at less than 0.5 standard deviation of the mean performance of students with mental retardation, as reported in Table 2. In Video 1, all target behaviors were performed within the normative range. In Video 2, verbal behaviors (i.e., rate and percentage of time initiating and responding) were performed below the normative range, whereas attending and engaging in distracting motor behavior were performed within the range. Conversely, in Video 3, verbal behaviors and engaging in distracting motor behavior were within the normative range, and attending was below the range. Finally, in Video 4, all behaviors were performed below the range, except engaging in distracting motor behavior, which was performed at the mean rate of the students with mental retardation. Systematically manipulating the range of behavioral dimensions allowed us to isolate the effects of specific behavioral components on judges' subjective evaluation of performance.

Each video scene was approximately 3 min long (range, 2.5 to 3.4 min) and depicted two actors eating at a round lunch table (1.3 m diameter) in a lunchroom similar to that of the target high school. The actors in the videos engaged in conversation according to the prepared scripts, systematically varying their target behaviors. At the same time, they maintained all other behaviors, such as positive affect and appropriateness of initiations, within the normative range observed for general education students. Conversational topics, such as peers or social events, were chosen from the list of the 10 topic areas. To insure the social validity of the language and topics of the scripted conversation among high school students, the four scripts were evaluated by 11 general and special education high school students (4 African Americans, 7 Caucasians; 9 females, 2 males). Based upon students' feedback, changes were made in the scripts, which then were reevaluated by the students. Students suggested changes in wording only (e.g., "That's cool" vs. "I think that's a great idea"), rather than content.

Two female college students (22 years old) majoring in special education, one Caucasian and one African-American, volunteered to participate as actors in the four scripted video scenes. Each was compensated with an honorarium of $50.00. The students were chosen based on their youthful appearance; similarity in grooming, dress, and expressive language to that of high school students; and lack of idiosyncratic or distracting behaviors (e.g., rolling one's eyes). To ensure the fidelity with which target behaviors were performed, two trained observers independently conducted continuous observation in real time of conversational interactions portrayed in the four videos according to the behavior definitions. Target behaviors were found to occur as indicated in the top panel of Table 3. Mean overall interobserver agreement calculated per behavior per video was 96% (range, 70% to 100%).

Comparison of Judges' Ratings to Behavioral Dimensions

We recruited 60 individuals who were not informed of the purpose of the study to judge the target behaviors portrayed in the videos. The judges represented six groups with whom the students with mental retardation were likely to interact: 10 general education high school students, 10 special education high school students, 10 general education high school teachers, 10 special education high school teachers, 10 employers, and 10 employees. Judges were recruited by asking for volunteers to participate in a social skills study. Informed consent was obtained from each judge. Members within each group were chosen to approximate the

ethnic and gender composition of students in the high school in which the target behaviors were derived: 20% African American females, 30% Caucasian females, 20% African American males, and 30% Caucasian males. Special education students were chosen from self-contained and resource room classes and were reported as having learning disabilities or mental retardation. General and special education students represented Grades 9 through 12 in proportion to the high school student population, with decreasing numbers of students in advancing grades.

General and special education teachers, employers, and employees were chosen in proportion to age groups in the local population: 40% were 22 to 36 years old and 60% were 37 or older. General education teachers taught courses within a range of content areas (e.g., science, social studies, English). Special education teachers were chosen from resource or self-contained rooms that served a range of students, including those with autism, mental retardation, learning disabilities, behavior disorders, and multiple handicaps. Employers and employees were chosen equally across professional, clerical, food service, and other entry-level jobs.

Each of the judges met individually with the second or third author for approximately 25 min to rate the video performances. Rating sessions were conducted in a small office in the high school for the students and teachers and in quiet locations convenient to their work sites for the employers and employees. First, an author read the following instructions to a judge, which were adapted from Quinn et al. (1992) and Sherman et al. (1992):

> I would like you to watch four videotaped situations involving 2 high school students having a conversation while eating in a school lunchroom. Each situation is about 3 minutes long. Please focus your attention on the person on the left of your screen and disregard the other person in the situation. After the scene, I will stop the tape and ask you seven questions about the conversational skills of the student. For the first five questions, you will circle the number on the answer sheet that best describes your opinion. You will rate the social skills of the person based on the following scale: 1 = *strongly disagree,* 2 = *mildly disagree,* 3 = *don't feel strongly either way,* 4 = *mildly agree,* 5 = *strongly agree.* [The five questions consisted of responding to the following statements: "The student has good conversational skills." "The student acts like most high school students when they eat lunch in a school cafeteria." "Most high school students would probably act like the student in this conversation." "The student's conversational behavior looks acceptable to me." "Most high school students would probably enjoy having a conversation with someone who acts like the student."] The last two questions will be short answer. In order to get as much information as possible about the student's conversational skills, I may also ask you some follow-up questions. [The two questions were "What did you like about how the student acted in this conversation?" and "What could the student have changed or added to improve how she acted in the conversation?"] Do you have any questions before we begin?

Next, the authors presented each of the four videotapes in random order on a video cassette player, pausing after each videotape for the judges to respond to the questions on a written questionnaire. The authors read all questions aloud, clarifying their meaning,

if necessary. For example, a definition would be provided if a judge asked what "conversational behavior" meant. The authors then circled the choices for the first five questions as verbalized by the judges. For the last two open-ended questions, the authors transcribed the judges' verbal responses, asking additional questions for clarification, as necessary. The process was repeated after each videotape was shown. At the close of the evaluation sessions, judges were given $3.00 in coupons for use at a local fast-food restaurant for their participation.

Data Analysis

Data analysis included descriptive and inferential statistical tests. To summarize findings, means and standard deviations were computed for responses to the five quantitative questions asked for each of the four videotapes. In addition, a mean score was obtained for all five questions for each video. To test for significant differences in judges' ratings on the four videos and significant effects of demographic variables of judges on video ratings, MANOVAs for repeated measures were conducted. We also calculated Kendall's coefficient of concordance to determine the extent of interjudge agreement. Finally, we used the constant-comparative method (Glaser & Strauss, 1967; Lincoln & Guba, 1985) to analyze qualitative responses to the two open-ended questions asked for each videotape. Using this method, responses were independently classified by three authors into provisional categories based on similarity of meaning. Classifications were then compared and revised, and, based on consensus among raters, definitions of categories and rules for inclusion were developed. Next, responses were independently reassigned to the revised categories, findings were compared, and agreement was computed. Mean agreement for assignment to content categories ($n = 7$) was 96% (range, 91% to 100% per question per video). Disagreements were reexamined and discussed until consensus on classification was reached.

RESULTS

Mean Ratings

Means and standard deviations for judges' responses to each of the five quantitative questions for the four videotapes are shown in the bottom section of Table 3. Low standard deviations indicate little variability across judges. The effects of varying levels of performance of target behaviors are illustrated by comparing the top and bottom sections of Table 3. When all six target behaviors were performed within the normative range (Video 1), judges' mean scores indicated strong agreement that the targeted actor in the video displayed desirable conversational skills (i.e., had good conversational skills, acted like a typical high school student, had acceptable conversational behavior, would be an enjoyable conversational partner for a high school student). Conversely, when all target behaviors were performed outside the normative range (Video 4), judges strongly agreed that the actor's conversational social behavior was unacceptable. These findings indicate that the target behaviors (e.g., initiating, responding, attending, not engaging in distracting motor behavior), when performed as a group, yielded favorable ratings by judges.

The effects of the two verbal behaviors—initiating and responding—in relation to the normative range are illustrated by the mean scores obtained for Videos 1 and 2. When verbal behaviors were performed within the normative range (Video 1), judges' scores indicated that performance was acceptable. When verbal behaviors were performed below the normative range and attending and engaging in distracting motor behaviors were within the range (Video 2), judges' ratings were similar to those in Video 4, for which judges agreed that performance was unac-

Table 4
Extension of Tukey's Multiple Range Test of Judges' Mean Scores on Four Videos

	M	Video 1 4.59	Video 2 1.50	Video 3 3.23	Video 4 1.32
Video 1	4.59				
Video 2	1.50	3.09*			
Video 3	3.23	1.36*	1.73*		
Video 4	1.32	3.27*	0.18	1.91*	

Note. HSD (honest significant difference) = .5101.
* $p < .05$.

ceptable. To isolate the effects on judges' ratings of attending behavior, Videos 1 and 3 may be compared. Attending within the normative range was associated with highly favorable ratings (Video 1). In contrast, when attending was below the normative range and verbal behavior and distracting motor behavior were within the range (Video 3), ratings were moderate (i.e., *do not feel strongly either way*). Thus, verbal behavior rather than attending appeared to be more critical to ratings of acceptability, although attending influenced ratings. The effects of distracting motor behavior cannot be isolated across the four videos because it was not manipulated independently; however, when performed in combination with little verbal or attending behavior, low ratings of acceptability were obtained.

To corroborate findings, a MANOVA for repeated measures was conducted that revealed significant differences among judges' mean ratings of the four videos (see Table 3). An extension of Tukey's multiple range test was conducted to determine the source of significance. Results indicated that five of the six paired comparisons differed significantly at the .05 level of confidence (see Table 4). Thus, verbal behaviors performed within the normative range (Videos 1 and 3, Table 3) were associated with significantly higher ratings by judges than verbal behaviors performed below the normative range (Videos 2 and 4, Table 3). In addition, a repeated measures mixed design MANOVA was conducted to assess the effects of demographic characteristics of judges (i.e., group membership [e.g., student, teacher, employer], ethnicity, gender) on ratings of videos. No significant effects were found for demographic variables on judges' mean video ratings (see Table 5).

Finally, to assess agreement among all 60 judges and within groups of judges on their overall rating of each video (i.e., sum of responses to questions), we calculated Kendall's coefficient of concordance (*W*; Siegel, 1956). Kendall's *W* expresses the degree of association among three or more variables measured in ranks and is useful in evaluating interjudge agreement. In using *W*, the degree of agreement among *k* judges is a function of the degree of variance among the sums of object ranks; the greater the variance, the less the concordance. *W*s can be tested for significance using a chi-square formula with the number of objects rated minus 1 serving as the degrees of freedom. A high or significant value of *W* (range, 0 to 1) indicates that judges are employing similar standards in rating objects.

In this analysis, we examined the overall ratings (sum of responses to questions) for

Table 5
MANOVA Analysis of Effects of Demographic Characteristics of Judges on Video Ratings

Demographic characteristic	*SS*	*df*	*MS*	*F*
Group Membership × Video	154.16	15	10.28	0.93
Ethnicity × Video	24.27	3	8.09	0.73
Gender × Video	54.19	3	18.06	1.63

Table 6
Sample and Subset Videotape Rankings and Kendall Coefficients of Concordance (*W*)

Groups	Cases	Video 1	Video 2	Video 3	Video 4	*W*	Chi-square[a]	Significance
Overall	60	4	2	3	1	.85	152.76	.0001
General education teachers	10	4	1.5	3	1.5	.84	25.09	.0001
Special education teachers	10	4	2	3	1	.95	28.45	.0001
General education students	10	4	1	3	2	.84	25.29	.0001
Special education students	10	4	2	3	1	.77	23.10	.0001
Employees	10	4	1	3	2	.95	28.45	.0001
Employers	10	4	2	3	1	.85	25.62	.0001

[a] $df = 3$.

each of the four videotapes ($df = 3$). For each judge, we ranked the videos based on the overall ratings and calculated Kendall's *W*. Then we tested the result using the chi-square formula recommended by Siegel (1956). As shown in Table 6, the overall *W* was .85, which was significant at the .0001 level. Subsequently, we analyzed each subset of the population to ensure that, despite heterogeneity of composition, concordance was present both within groups of judges and across judges. The resulting *W*s and their significance are shown in Table 6. These results show that judges, whether evaluated as a whole or as membership groups, displayed considerable agreement, indicating that they employed similar criteria in the evaluation of the four videotaped performances. Specifically, the subgroups were consistent in their rankings of the best (Video 1; mean rank = 4) and second best (Video 3; mean rank = 3) performances. (A value of 4 indicated the highest ranking and 1 indicated the lowest.) All subgroups ranked Video 2 (mean rank = 1.58) and Video 4 (mean rank = 1.42) lower, although there was some disagreement regarding which of the two was the least favorable.

Qualitative Responses

As suggested by Quinn et al. (1992), we asked judges two open-ended questions—what they liked about the videotaped performance and what needed to be improved—to determine whether their verbal comments were consistent with their ratings of target behaviors. Responses to each of the two questions obtained from all 60 judges for each of the four videos resulted in a total of 875 codable comments (available on request). For example, when asked, "What did you like about how the student acted?" comments included "She was responsive to the questions that were asked," "She initiated conversation frequently," and "She seemed interested in her friend." The numbers of judges who provided a comment that corresponded to the target behaviors addressed in this study are found in Table 7. Findings corroborated variations in judges' ratings when behavioral components were performed within or outside the normative range. High frequencies of positive comments were provided when target behaviors were performed within the normative range; conversely, many suggestions for improvement were associated with performance outside that range. The high frequency of comments that corresponded to target behaviors was particularly noteworthy because judges were not instructed to attend to these behaviors. Judges also provided comments related to the qualitative aspects of the actors' conversational behavior that we controlled for, such as their positive affect, appropriateness of initiations, and choice of topics.

Table 7
Number of Judges Who Provided Comments that Corresponded to Target Behaviors

Behavioral component	Performed within normative range		Performed outside normative range	
	Liked	Need to improve	Liked	Need to improve
Initiating and responding	44 (73)[a]	2 (3)	5 (8)	49 (82)
Attending	34 (57)	0 (0)	4 (7)	30 (50)
Engaging or not engaging in distracting motor behavior	8 (13)	1 (2)	3 (5)	44 (73)

[a] Percentage of total number of judges ($N = 60$) in parentheses.

DISCUSSION

Using a four-step social validation process—nominating target behaviors, establishing a normative range of performance, manipulating simulations along selected behavioral dimensions, and comparing judges' ratings to performance along those behavioral dimensions—we successfully identified and validated potentially critical skill components that constitute high school students' conversational behavior. Multiple quantitative and qualitative measures corroborated findings that the simulated performance of targeted components of conversational skills (i.e., rate and percentage of time initiating and responding, percentage of time attending and not engaging in distracting motor behavior) related to favorable ratings by a wide variety of judges.

This study makes at least three contributions to the literature on social skills. First, the investigation demonstrated the utility of a four-step process for identifying and validating social skill components that relate to favorable judgments by others. We empirically demonstrated the practicality of a method for identifying frequencies of simulated performance of targeted skills that relate to ratings of social acceptability by others. In addition, this process allowed us to evaluate the effects of simulations of the target behaviors when performed individually and in combination with each other. These findings are important because there is limited knowledge with respect to social behaviors that may relate to favorable outcomes, such as increases in social acceptance and social interaction (Chadsey-Rusch, 1992; Odom & McConnell, 1992; Siperstein, 1992). Identifying skills that relate to favorable judgments by others suggests potential targets for social skills interventions.

Second, the four-step process that we developed addressed several limitations of previous attempts to validate social behaviors (Minkin et al., 1976; Quinn et al., 1992; Sherman et al., 1992). Specifically, Quinn et al. and Sherman et al. targeted behaviors based on face validity alone, as derived from the literature; Minkin et al. based behaviors on clinical samples of conversation. In contrast, we used social comparison methods by observing high school students in the actual setting, comparing conversational behavior of students with and without moderate or severe mental retardation, and establishing a normative range of identified behaviors based on a sample of behavior. Although students were aware that they were being observed, they were unaware of the behaviors of interest. In addition, we sampled a wider range of judges (e.g., ethnically diverse) than in previous studies and included representatives of the youths themselves and their peers without disabilities, as suggested by Minkin et al. Despite the diversity of judges, greater agreement was found than in previ-

ous studies, suggesting that identified behaviors were universally valued across groups.

Further, in our study, the actual frequencies of target behaviors performed by the actors in the videotapes were empirically controlled, unlike in the Minkin et al. (1976) or Sherman et al. (1992) studies. Specifically, behaviors were performed within either 0.5 or 1 standard deviation of mean performance of students with or without mental retardation, respectively. Finally, we chose to investigate everyday conversational social skills of high school students rather than task-related social skills of adults. In developing instructional curricula, it is important to target skills that are used and reinforced frequently in the everyday environment. Because postsecondary adult outcomes for students with disabilities typically are poor (e.g., economic dependence, segregation, unemployment; Harris & Associates, 1994; Wagner, 1995), it is imperative that practitioners begin to address critical skills while students are still in high school (Moore et al., 1990).

Third, applied behavior analysis traditionally has targeted objectively measured behavior (Minkin et al., 1976; Quinn et al., 1992). As behavior analysts begin to address behaviors such as maintaining social relationships or developing social networks, multiple measures of social validity, such as those conducted in this study, will become more essential (Horner, 1996; Minkin et al., 1976).

Several limitations of this study are noteworthy. First, subtle and influential features of social interaction were not monitored and may have been confounded with the quantitative manipulations. Examples of these features include the type of initiations and responses (open-ended vs. yes-no, obligatory vs. nonobligatory, questions vs. comments), their timing (do they interrupt each others' turns, are the latencies between turns too long), and the types of nonverbal behavior accompanying the initiations and responses. Videotaping participants' social interactions would have allowed us to rescore our observations as needed, taking into account multiple variables that we could not code reliably when scoring social interaction in the natural setting.

Second, our efforts leave unaddressed the question of whether performing the targeted social behaviors is functionally related to favorable outcomes for students, such as having peers choose to sit with them at lunch or having a best friend in a general education class. Further, we did not demonstrate whether students with moderate or severe mental retardation can learn to perform the targeted skills, although our previous work (e.g., Hughes et al., 1996) indicated that they can. Whereas some may argue that the ultimate test is the functional validation of targeted behaviors—whether performing the skills results in favorable outcomes—that is a terminal issue (Hawkins, 1991).

Future research should use experimental analysis methods to address the functional validity of the skills identified in this study to determine whether their performance relates to increased social acceptance and social interactions in students' everyday lives (Hawkins, 1991). Further, whereas performance of the targeted skills may be functional for a subset of high school students with moderate or severe mental retardation, it must be acknowledged that they may be inappropriate instructional targets for some of this population who, for example, may lack prerequisite skills or may have already mastered the skills.

A related issue is that generality of findings may be limited by the fact that only two actors were depicted across videotapes. Therefore, it is possible that judges' differential ratings were influenced by personal characteristics of the actors rather than variations in rates of target behaviors. Finally, our manipulations of target behaviors failed

to isolate the effects of some potentially critical skill components. For example, the influence of distracting motor behavior on judges' perceptions was not investigated in isolation. In addition, verbal behaviors were manipulated as a response class, rather than investigating the effects of initiating or responding separately. Future research should provide a more fine-grained analysis of the individual skill components that affected judges' ratings.

In summary, we conducted a multiphase study to identify and validate potentially critical conversational social skills among high school students. Whereas we demonstrated the social validity of critical skill components, the next logical step would be to investigate the functional validity of the targeted behaviors. Future research should address developing interventions to teach the potentially critical skills identified and validated by the four-step process. Researchers should then evaluate the effects of target skill performance on the social interactions and social relationships of high school students with and without mental retardation.

The procedures used in this investigation, in which measures of performance were compared with subjective evaluations by important others, represent a type of research—social validation—that began with Wolf (1978) and has continued for over 20 years in the *Journal of Applied Behavior Analysis*. Social validation research addresses the social importance of the goals, procedures, and effects of intervention efforts. Because applied behavior analysis focuses on applied problems of human behavior, the social significance and acceptability of an intervention must be assessed. However, since 1978, behavior analysts have struggled to develop methods for assessing social perception of behavior-change efforts. This study extends the growing technology of social validation assessment (e.g., Hawkins, 1991; Sherman et al., 1992; Van Houten, 1979; Wolf, 1978) and offers another tool for identifying the variables that influence human behavior.

REFERENCES

Bakeman, R., & Gottman, J. A. (1986). *Observing interaction: An introduction to sequential analysis.* New York: Cambridge University Press.

Breen, C. G., Kennedy, C. H., & Haring, T. G. (Eds.). (1991). *Social context research project: Methods for facilitating the inclusion of students with disabilities in integrated school and community contexts.* Santa Barbara, CA: University of California.

Brown, L., Branston, M. B., Hamre-Nietupski, S., Pumpian, I., Certo, N., & Gruenewald, L. (1979). A strategy for developing chronological-age-appropriate and functional curricular content for severely handicapped adolescents and young adults. *Journal of Special Education, 13,* 81–90.

Butterworth, J., Jr., & Strauch, J. D. (1994). The relationship between social competence and success in the competitive work place for persons with mental retardation. *Education and Training in Mental Retardation and Developmental Disabilities, 29,* 118–133.

Chadsey-Rusch, J. (1992). Toward defining and measuring social skills in employment settings. *American Journal on Mental Retardation, 96,* 405–418.

Chadsey-Rusch, J., Gonzalez, P., Tines, J., & Johnson, J. R. (1989). Social ecology of the workplace: Contextual variables affecting social interactions among employees with and without mental retardation. *American Journal on Mental Retardation, 94,* 141–151.

Clark, G. M., Field, S., Patton, J. R., Brolin, D. E., & Sitlington, P. L. (1994). Life skills instruction: A necessary component for all students with disabilities. A position statement of the Division on Career Development and Transition. *Career Development for Exceptional Individuals, 17,* 125–134.

Fawcett, S. B. (1991). Social validity: A note on methodology. *Journal of Applied Behavior Analysis, 24,* 235–239.

Ferguson, B., McDonnell, J., & Drew, C. (1993). Type and frequency of social interaction among workers with and without mental retardation. *American Journal on Mental Retardation, 97,* 530–540.

Fey, M. E. (1986). *Language intervention with young children.* San Diego, CA: College-Hill Press.

Foster, S. L., & Cone, J. D. (1986). Design and use of direct observation procedures. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment* (2nd ed., pp. 253–324). New York: Wiley.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory.* Chicago: Aldine.

Goldstein, H., Kaczmarek, L., Pennington, R., & Sha-

fer, K. (1992). Peer-mediated intervention: Attending to, commenting on, and acknowledging the behavior of preschoolers with autism. *Journal of Applied Behavior Analysis, 25,* 289–305.

Halpern, A. S. (1992). Transition: Old wine in new bottles. *Exceptional Children, 58,* 202–211.

Haring, T. G. (1991). Social relationships. In L. H. Meyer, C. A. Peck, & L. Brown (Eds.), *Critical issues in the lives of people with severe disabilities* (pp. 195–217). Baltimore: Paul H. Brookes.

Harris and Associates, Inc. (1994). *N.O.D./Harris survey of Americans with disabilities.* New York: Author.

Hawkins, R. P. (1991). Is social validity what we are interested in? Argument for a functional approach. *Journal of Applied Behavior Analysis, 24,* 205–214.

Horner, R. H. (1996, May). *Current issues in supported employment.* Paper presented at the meeting of the Association for Behavior Analysis, San Francisco.

Hughes, C., Hwang, B., Kim, J. H., Eisenman, L. T., & Killian, D. J. (1995). Quality of life in applied research: A review and analysis of empirical measures. *American Journal on Mental Retardation, 99,* 623–641.

Hughes, C., Killian, D. J., & Fischer, G. M. (1996). Validation and assessment of a conversational interaction intervention. *American Journal on Mental Retardation, 100,* 493–509.

Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification, 1,* 427–452.

Kazdin, A. E. (1982). *Single case research design: Methods for clinical and applied settings.* New York: Oxford University Press.

Koegel, R. L., & Frea, W. D. (1993). Treatment of social behavior in autism through the modification of pivotal social skills. *Journal of Applied Behavior Analysis, 26,* 369–377.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.

Minkin, N., Braukman, C. J., Minkin, B. L., Timbers, G. D., Timbers, B. J., Fixsen, D. L., Phillips, E. L., & Wolf, M. M. (1976). The social validation and training of conversational skills. *Journal of Applied Behavior Analysis, 9,* 127–139.

Moore, S. C., Agran, M., & McSweyn, C. A. (1990). Career education: Are we starting early enough? *Career Development for Exceptional Individuals, 13,* 129–134.

Nelson, R. O., & Hayes, S. C. (1979). The nature of behavioral assessment. *Journal of Applied Behavior Analysis, 12,* 491–500.

Odom, S. L., & McConnell, S. R. (1992). Improving social competence: An applied behavior analysis perspective. *Journal of Applied Behavior Analysis, 25,* 239–243.

Prutting, C. A., & Kirchner, D. M. (1987). A clinical appraisal of the pragmatic aspects of language. *Journal of Speech and Hearing Disorders, 52,* 105–119.

Quinn, J. M., Sherman, J. A., Sheldon, J. B., Quinn, L. M., & Harchik, A. E. (1992). Social validation of component behaviors of following instructions, accepting criticism, and negotiating. *Journal of Applied Behavior Analysis, 25,* 401–413.

Rusch, F. R., DeStefano, L., Chadsey-Rusch, J., Phelps, L. A., & Szymanski, E. (Eds.). (1992). *Transition from school to adult life: Models, linkages, and policy.* Sycamore, IL: Sycamore.

Schalock, R. L. (1986). Employment outcomes from secondary school programs. *Remedial and Special Education, 7,* 37–39.

Schwartz, I. S., & Baer, D. M. (1991). Social validity assessments: Is current practice state of the art? *Journal of Applied Behavior Analysis, 24,* 189–204.

Sherman, J. A., Sheldon, J. B., Harchik, A. A., Edwards, K., & Quinn, J. M. (1992). Social evaluation of behaviors comprising three social skills and a comparison of the performance of people with and without mental retardation. *American Journal on Mental Retardation, 96,* 419–431.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Siperstein, G. N. (1992). Social competence: An important construct in mental retardation. *American Journal on Mental Retardation, 96,* iii–vi.

Storey, K., & Horner, R. H. (1991). Social interactions in three supported employment options: A comparative analysis. *Journal of Applied Behavior Analysis, 24,* 349–360.

Tapp, J., Wehby, J. H., & Ellis, D. N. (1995). A multiple option observation system for experimental studies: MOOSES. *Behavior Research Methods, Instruments, and Computers, 27,* 25–31.

Van Houten, R. (1979). Social validation: The evolution of standards of competency for target behaviors. *Journal of Applied Behavior Analysis, 12,* 581–591.

Wagner, M. (1995). *Transition from high school to employment and post-secondary education: Interdisciplinary implications for youths with mental retardation.* Paper presented at the 119th annual meeting of the American Association on Mental Retardation, San Francisco.

Warren, S. F., Rogers-Warren, A., & Baer, D. M. (1976). The role of offer rates in controlling sharing by young children. *Journal of Applied Behavior Analysis, 9,* 491–497.

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11,* 203–214.

Received July 5, 1996
Initial editorial decision September 9, 1996
Final acceptance April 21, 1998
Action Editor, James W. Halle

STUDY QUESTIONS

1. Describe the four-step sequence proposed by the authors for identifying and validating critical components of social skills.

2. What procedures did the authors use to select target behaviors? What was the outcome of this process?

3. Who was selected for more detailed observations, and why was the performance of these groups compared?

4. Briefly describe the details of the observation procedure used during the group comparison.

5. In what ways were performances of the two groups similar and different?

6. How was the normative range of performance identified, manipulated, and evaluated?

7. What dimensions of social interaction appeared to affect the ratings?

8. One conclusion based on results of this study is that more normalized interactions might result from increasing certain aspects of social behavior in students with mental retardation. However, the results were derived from a social comparison based on quantitative characteristics of behavior. Under what conditions might these results not generalize to a target group selected for intervention?

Questions prepared by Juliet Conners and Eileen Roscoe, The University of Florida